Monte Carlo Sampling Methods Homework 4

Zihan (Steven) Zhang

November 17, 2022

I have consulted and discussed the materials with Qianyu Zhu.

1 Exercise 39

In the simplest realization of the partial resampling principle, the transition of the Markov chain is drawn directly and independently from a conditional distribution:

$$p(t+1, y_{j_t}|t, x) = \pi(y_{j_t}) \tag{1}$$

Markov chains of this type are called *Gibbs samplers*, in which samples are drawn independently from a conditional distribution of π. However, it is not applicable unless the coordinates can be chosen so that the conditional distribution of each component is very simple, like Gaussian. In this question, we aim to use a Gibbs sampler to generate samples of the Ising model and plot the histogram of the values of the magnetization:

$$f(\sigma) = \sum_{\vec{i} \in \mathbb{Z}_L^2} \sigma_{\vec{i}} \tag{2}$$

Also, following the lecture note, if we consider a vector σ indexed on the periodic 2-dimensional lattice $\mathbb{Z}_L^2$ and with values in $\{-1, 1\}$. If we assign the density:

$$\pi(\sigma) = \frac{e^{\beta \sum_{\vec{i} \leftrightarrow \vec{j}} \sigma_{\vec{i}} \sigma_{\vec{j}}}}{\mathcal{Z}} \tag{3}$$

to the σ variables, this becomes the Ising model of statistical physics. Here $\vec{i}, \vec{j} \in \mathbb{Z}_L^2$ and $\vec{i} \leftrightarrow \vec{j}$ indicate that $\vec{i}$ and $\vec{j}$ are neighboring sites on the lattice. The constant $\beta > 0$ is related to a physical temperature via $k_B T = \beta^{-1}$ where k_B is the Boltzmann constant. Thus:

$$\pi(\sigma_{\vec{i}_t}|\sigma_{[\vec{i}_t]}) = \omega_+ \delta(\sigma_{\vec{i}_k} - 1) + \omega_- \delta(\sigma_{\vec{i}_k} + 1). \tag{4}$$

where:

$$w_+ = \frac{e^{\beta \sum_{\vec{i} \leftrightarrow \vec{j}} \sigma_{\vec{i}}}}{e^{\beta \sum_{\vec{i} \leftrightarrow \vec{i}_t} \sigma_{\vec{i}}} + e^{-\beta \sum_{\vec{i} \leftrightarrow \vec{i}_t} \sigma_{\vec{i}}}} \tag{5}$$

and:

$$w_- = \frac{e^{-\beta \sum_{\vec{i} \leftrightarrow \vec{i}_t} \sigma_{\vec{i}}}}{e^{\beta \sum_{\vec{i} \leftrightarrow \vec{i}_t} \sigma_{\vec{i}}} + e^{-\beta \sum_{\vec{i} \leftrightarrow \vec{i}_t} \sigma_{\vec{i}}}} \tag{6}$$

We take $L = 10, \beta = 0.1$, and the maximum number of iterations is 10^6 for both the random and ordered flip positions and compute the integrated autocorrelation time for the magnetization. In Figure 1, we respectively select i randomly or sweep through the lattice deterministically. Both results of the magnetization of lattice approximated are like Gaussian random variables. However, the results are only reliable when $\beta < 0.441$.

Firstly, we fix other parameters and use different lattice sizes from 5 to 30. Too large L would require a huge amount of iterations and exponentially increase the computation costs. Figure 2 describes the quadratic relationship between autocorrelation time and lattice dimension L for both randomized and structured Gibbs. The structured Gibbs has a faster convergence speed by a nearly constant factor.

Secondly, we fix other parameters and change the temperature. Figure 3 shows the inverse relationship between temperature T and ACT when $\beta < 0.441$. Similarly, the structured Gibbs sampler converges faster than the randomized version. Also, as β gradually grows, ACT would firstly increase, decrease, and then oscillate. Theoretically, ACT should increase exponentially with β. As previously mentioned, when $\beta > 0.441$, the results become unreliable since it becomes hard for

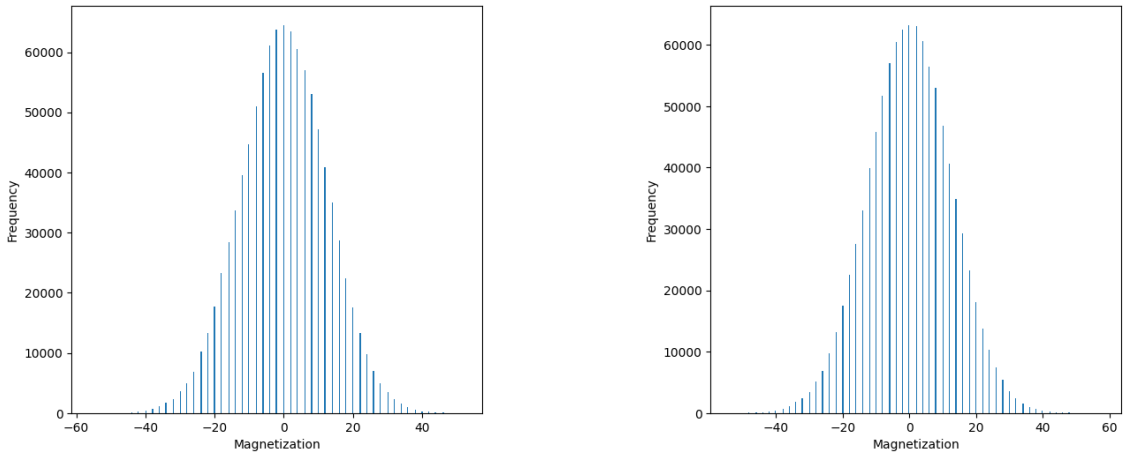

Figure 1: Examples of randomized Gibbs sampler (left) and structured Gibbs sampler (right).

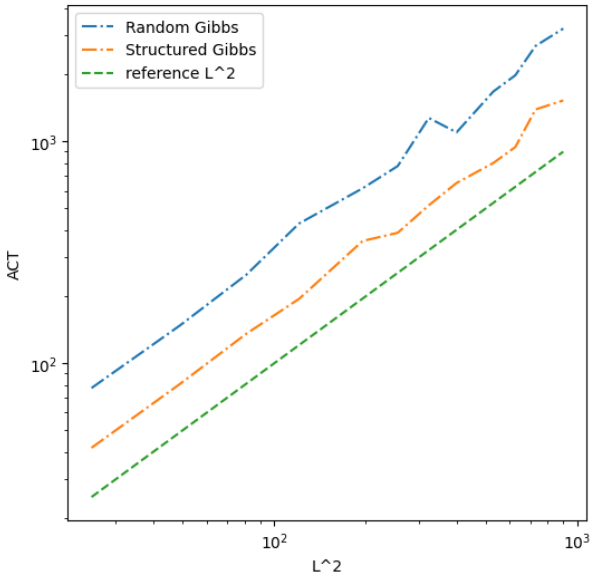

Figure 2: The quadratic relationship between autocorrelation time (ACT) and lattice dimension L.

the chain to jump from one magnetization into another, as shown in Figure 4. When $\beta = 0.1$, the magnetization is evenly distributed around 0. When $\beta = 0.46$, the magnetization is always negative and changes in the sign very rarely when β is beyond the threshold, which would lead to a longer ACT time. Thus, when β increases and the chain's length is relatively short, the magnetization would have a smaller variance until the magnetization change sign.

2 Exercise 42

From Example 22 in the lecture note, considering the Ising model and suppose we are constructing a chain $X^{(k)}$ with values in the set of $L \times L$ matrices with entries $\{-1, 1\}$, and that, at each step of the chain $Y^{(k+1)}$ corresponds to flipping the sign of a single entry of $X^{(k)}$. Suppose that the index of the spin at which a sign flip is proposed at step k is $\vec{i}_k$, then:

$$\frac{\pi\left(Y^{(k+1)}\right)}{\pi\left(X^{(k)}\right)} = e^{-4\beta X_{\vec{i}_k}(k)\sum_{\vec{j}\leftrightarrow\vec{i}_k} X_{\vec{j}}(k)} \tag{7}$$

Thus, we need only sum the spins of the neighbors of $\vec{i}_k$, an operation much less costly than the $\mathcal{O}(L^2)$ operations required to evaluate π itself (ignoring the normalization constant which we do not need to know).
Similar to Example 22, if the proposal is flipping the sign at $A[\vec{i}, \vec{j}]$, to achieve detailed balance, the acceptance rate is:

$$\min\left\{1, \frac{\pi(Y^{(t+1)})}{\pi(X^{(t)})}\right\} = \min\left\{1, e^{-2\beta X_{\vec{i}_k}(k)\sum_{\vec{j}\leftrightarrow\vec{i}_k} X_j(k)}\right\} \tag{8}$$

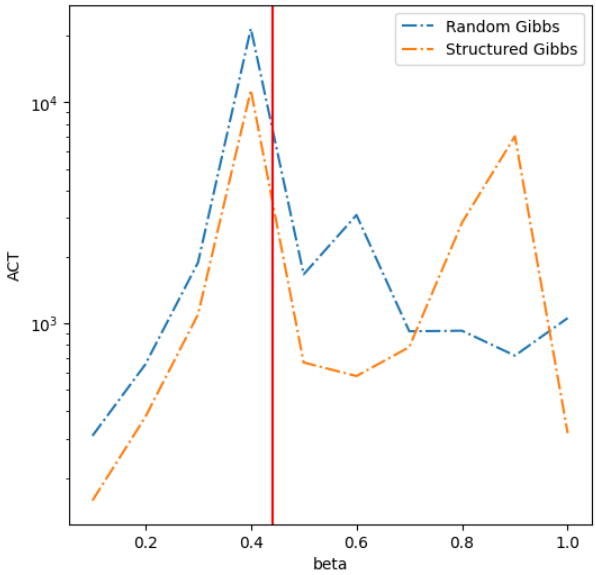

Figure 3: For $\beta < 0.441$ (red vertical line), the autocorrelation time increases with β, scaled inverse of temperature.

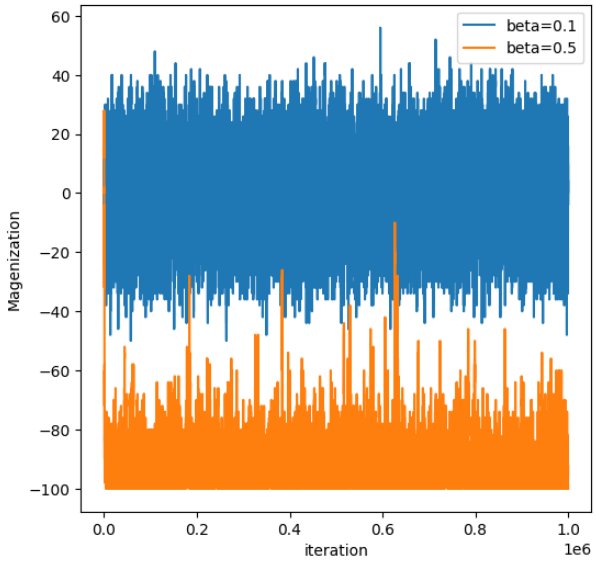

Figure 4: Relationship between magnetization and iteration when $\beta = 0.1$ and $= 0.5$ respectively.

When we fix the lattice size $L = 10$ and the maximum number of iterations is 10^6 and change β. ACT time would grow exponentially with β, consistent with the Gibbs sampler, as shown in Figure 5. If we fix $\beta = 0.1$, there is a quadratic relationship between L and ACT, also consistent with the Gibbs sampler, as shown in Figure 6. Moreover, the ACT for the Metropolis-Hasting algorithm is much smaller than that for randomized & structured Gibbs samplers.

3 Exercise 43

Referred to Section 4.7 in the lecture notes, we follow the procedures of Important MCMC sampling: Firstly generate N independent samples with equal initial weights and $\beta = 0$, as unbiased Bernoulli random variables, then for each k, we could design $\mathcal{T}_k$ such that $\pi_{k-1}\mathcal{T}_k = \pi_{k-1} \propto \pi^{(k-1)/N}\tilde{\pi}^{1-(k-1)/N} \propto \pi^{(k-1)/N}$. For Gibbs sampler with β, $\mathcal{T}_k$ is the original transition matrix with $\beta_k = \frac{k-1}{N}\beta$. For Metropolis-Hasting with β, the acceptance rate is $p_k = p^{(k-1)/N}$ and the proposal is symmetric, namely $\beta_k = \frac{k-1}{N}\beta$!. We could then use that $\mathcal{T}_k$ to update every sample in the ensemble. We use the Metropolis Hasting method to update every single sample, meaning that we propose a flip for this sample and accept the proposal with $p = \min\left(1, \frac{\pi_{k-1}(y)}{\pi_{k-1}(x)}\right)$. The weight ω will be used in the next step as $\frac{\pi_k(x)}{\pi_{k-1}(x)}$. To have a better convergence rate, at each kth step, we would make L^2 consecutive proposals for each sample. Thus, to update

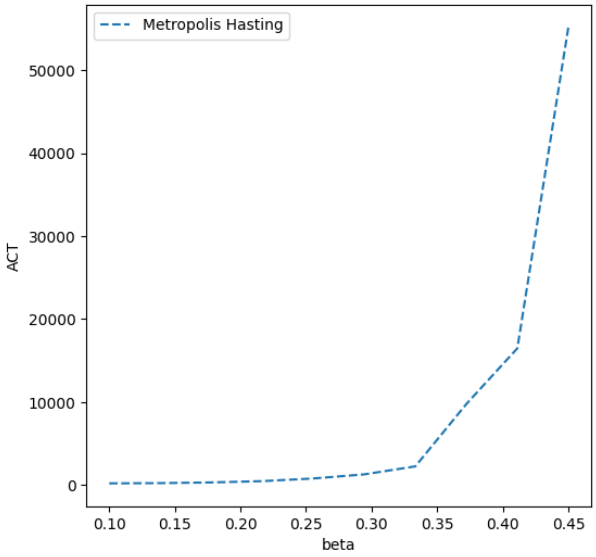

Figure 5: Relationship of β and ACT when sampling the 2d-Ising model.

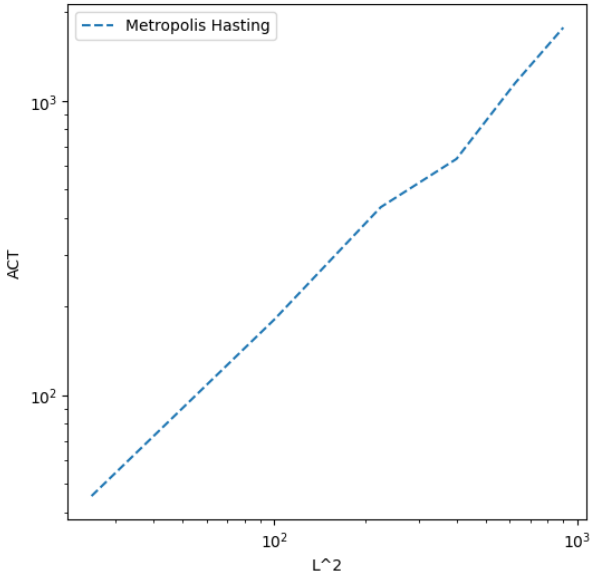

Figure 6: Relationship of L and ACT when sampling the 2d-Ising model.

each weight, we have $W^k \propto W^{k-1}\frac{\pi_k(x)}{\pi_{k-1}(x)} = W^{k-1}(x)\pi^{1/N}\tilde{\pi}^{-1/N} \propto W^{k-1}(x)\pi^{1/N}(x)$. In terms of resampling, we use the systematic resampling technique here. Finally, we could calculate the final magnetization using W^N and the final ensemble with a smaller bias.

Here we use 50 samples and 100 trials to calculate the variance with different values of N from 1 to 15, $L = 5, \beta = 0.2$. The parameters are chosen to make the computational cost not too expensive. Figure 7 shows the results with and without resampling, where the statistical variance decays with N increases. Also, after resampling, the variances largely decrease.

To compare Jarzynski's method and MH or Gibbs sampler, we first declare that Jarzynski's method is unbiased. When k is large, Jarzynski's method with $\pi_k = (\frac{\pi}{\tilde{\pi}})\tilde{\pi}$ will create $W^{(N)}$ such that $\log(W^{(N)}) \to \int V(x)\pi(dx)$, so the variance would vanish when $k \to \infty$. Still, the variance might be large and dominate the error when k is small. However, partial resampling with MH or Gibbs is biased if starting with a sample that is not generated from the target distribution. A longer chain would average out the bias, but when the dimension of lattice (L) or β is large, ACT would blow up, which leads to computational difficulty. In other words, we should use partial resampling with Metropolis-Hasting or Gibbs sampler for a small-scale problem since the bias would decrease when using a longer chain. However, for a large-scale problem, we should implement Jarzynski's method with resmpling which would have a small bias.

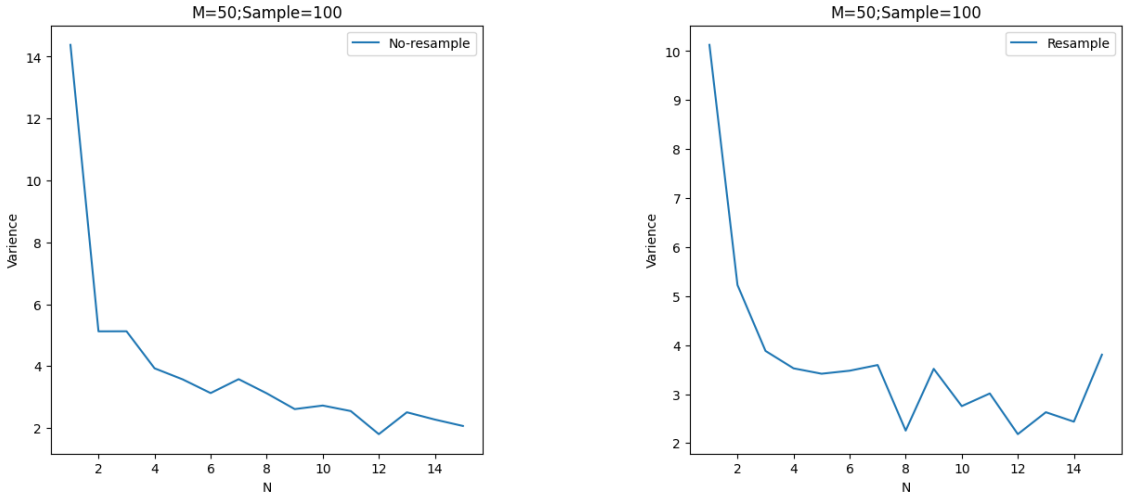

Figure 7: Jarzynski's method using Metropolis-Hasting sampling without (left) and with (right) resampling at each step.